The additivity of substrate fragments in enzyme–ligand binding

Thomas J Stout†, Carleton R Sage† and Robert M Stroud*

Background: Enzymes have evolved to recognise their target substrates with exquisite selectivity and specificity. Whether fragments of the substrate – perhaps never available to the evolving enzyme – are bound in the same manner as the parent substrate addresses the fundamental basis of specificity. An understanding of the relative contributions of individual portions of ligand molecules to the enzyme-binding interaction may offer considerable insight into the principles of substrate recognition.

Results: We report 12 crystal structures of *Escherichia coli* thymidylate synthase in complexes with available fragments of the substrate (dUMP), both with and without the presence of a cofactor analogue. The structures display considerable fidelity of binding mode and interactions. These complexes reveal several interesting features: the cofactor analogue enhances the localisation of substrate and substrate fragments near the reactive thiol; the ribose moiety reduces local disorder through additional specific enzyme–ligand interactions; the pyrimidine has multiple roles, ranging from stereospecificity to mechanistic competence; and the glycosidic linkage has an important role in the formation of a covalent attachment between substrate and enzyme.

Conclusions: The requirements of ligand–protein binding can be understood in terms of the binding of separate fragments of the ligand. Fragments which are subsystems of the natural substrate for the enzyme confer specific contributions to the binding affinity, orientation or electrostatics of the enzymatic mechanism. This ligand-binding analysis provides a complementary method to the more prevalent approaches utilising site-directed mutagenesis. In addition, these observations suggest a modular approach for rational drug design utilising chemical fragments.

Address: Departments of Biochemistry and Biophysics, School of Medicine, University of California San Francisco, San Francisco, California 94143-0448, USA.

†Present address: MetaXen, LLC, 280 E. Grand Avenue, South San Francisco, California 94080, USA.

*Corresponding author.
E-mail: stroud@msg.ucsf.edu

Key words: drug design, modular ligand binding, substrate fragments, substrate recognition, X-ray crystallography

Received: **20 March 1998**
Revisions requested: **20 April 1998**
Revisions received: **6 May 1998**
Accepted: **19 May 1998**

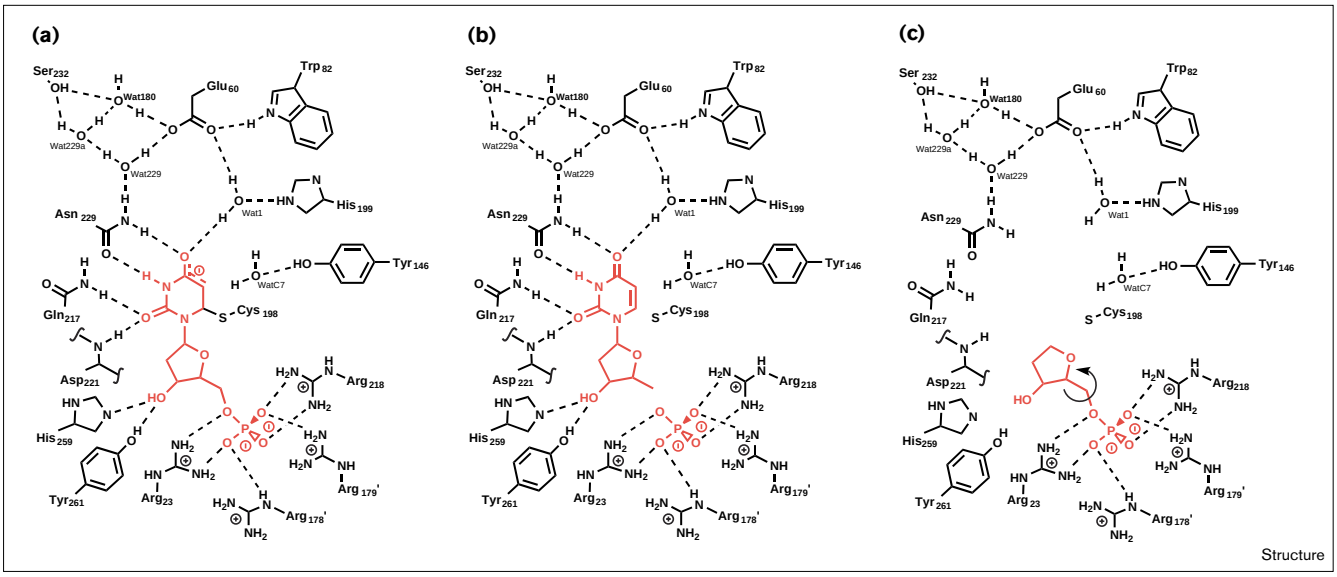
Structure 15 July 1998, **6**:839–848
http://biomednet.com/elecref/0969212600600839

Introduction

One of the fundamentals of life at the molecular level is that enzymes are required to act upon their substrates with speed, precision and fidelity. As a result, evolutionary pressures have caused these proteins to develop recognition systems for their target substrates which function with exquisite selectivity and specificity. Studies of the interactions between enzymes and their substrates have generally found that multiple interactions are typically generated within the enzyme around the available 'hooks' presented by the substrate. This multidentate approach to ligand binding provides the enzymatic system with both selectivity (improving the odds that the binding of similar but slightly different ligands will be disfavoured), and precision (lining up the substrate in a mechanistically competent orientation). The question of whether fragments of the substrate — individually perhaps never available to the evolving enzyme and thus not a component of the selective pressures of evolution — are bound in the same manner as the parent substrate addresses the fundamental basis of enzyme specificity. If additivity of components, in both energetic and structural terms, is realised, considerable insight may be gained into the principles of substrate recognition as well as providing a contribution to combinatorial approaches to drug design.

Because production of dTMP is a requirement for DNA synthesis and therefore cell division, thymidylate synthase (TS), the only biosynthetic source of cellular dTMP, has been extensively studied both as a drug target and as a model enzymatic system. Early studies of the catalytic mechanism of TS using organic chemical analogues proposed that the key step of the TS reaction involved the formation of a covalent bond between the substrate and a critical cysteine residue [1]. Investigations using molecular biology to alter TS residue by residue allowed an analysis of which amino acid residues were important in the TS reaction [1]. The determination of various TS crystal structures alone and in complexes with substrates and substrate analogues [2–6] further focused studies of the interactions between TS and its ligands. Combination of the latter two approaches has identified no fewer than 13 residues that interact directly, via hydrogen

Figure 1



An extensive hydrogen-bond network of interactions is formed between TS and the substrate dUMP. Substrate fragments utilise subsets of this same network. **(a)** In the presence of a cofactor analogue (not shown), dUMP forms a covalent adduct with the catalytic cysteine residue of TS. **(b)** The complex of the substrate fragments ddUrd and PO_4^{2-} with TS in the presence of a cofactor analogue (not shown) does not perturb any of these interactions but does not form the covalent adduct with the enzyme. Relative to the natural substrate, dUMP, this is equivalent to removal of the glycosidic bond. **(c)** The complexes of the substrate fragment phosphoribose (PR) with TS both with and without the cofactor analogue display rotational disorder about the glycosidic linkage. Thus, the specific interactions between the ribose moiety and the enzyme are either greatly reduced or eliminated, relative to the natural substrate complex, by the 'loss' of the pyrimidine ring.

bonding or hydrophobic interactions, or indirectly, via water molecules, with the substrate dUMP. The resulting network of contacts (Figure 1) illustrates the complexity of the interaction between TS and dUMP.

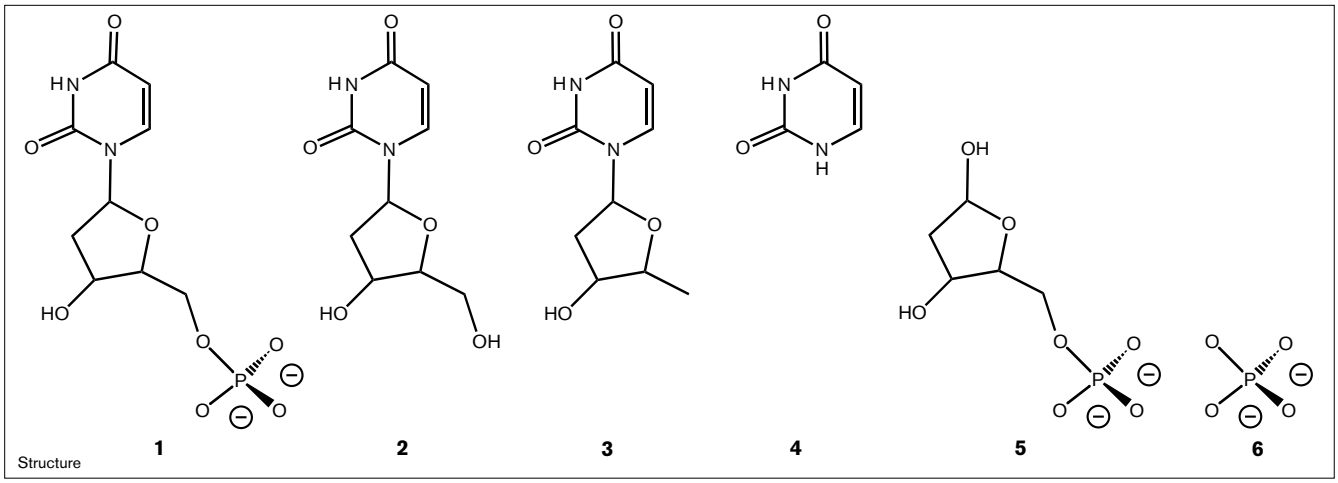
Many single- and multiple-site mutants of TS have been studied, both kinetically and structurally (reviewed in [1]), to address the binding of ligands and the catalytic mechanism of the enzyme. The atomic-resolution structures of TS and its mutants implicate these residues in substrate-binding affinity, orientation and specificity. Enzymes show considerable plasticity of mutant compensation and ligand binding, however [7]. In addition, very few residues are completely essential to the TS reaction (reviewed in [1]). Most problems associated with the use of site-directed mutagenesis to analyse an enzyme reaction are due to the inability to use genetic means to make small chemical changes at a particular position in the protein as, with few exceptions, the 20 naturally occurring amino acids are very different from each other. As a consequence, most amino acid sequence changes result in relatively large changes in molecular volume and local electrostatics, adding to the complexity of interpreting the chemical and structural rearrangements of the protein; the determination of the consequences of these changes on the enzyme reaction mechanism is also complicated.

In this study, we analyse the TS reaction — both biochemically and structurally — by systematically removing functional moieties of the substrate, dUMP. This was planned as an inversion of the usual process of dissecting enzymatic reactions by looking, not at modulations in the protein and their effects on the ligands through site-directed mutagenesis, but rather at incremental modifications in the bound ligands and their effect on the mode of binding. Complexes were studied with and without the cofactor analogue CB3717, and a range of dUMP 'fragments' (Figure 2). These fragments included intact 2′-deoxyuridine monophosphate (dUMP) (**1**), 2′-deoxyuridine (dUrd) (**2**), 2′,5′-dideoxyuridine (ddUrd) (**3**), uridine (Urd) (**4**), phosphoribose (PR) (**5**), and phosphate (PO_4^{2-}) (**6**). In each case, the composition of the protein scaffold remained constant and any changes in interactions can be interpreted as important aspects of the binding of the fragments. This approach allows the analysis of changes in enzyme–substrate interactions without altering the enzyme itself. The results complement and extend site-directed mutagenesis studies to help pinpoint the shared roles of substrates, substrate analogues and the protein in enzymatic reactions.

Results and discussion

The substrate and fragments studied are described in the Materials and methods section and are summarised in Figure 2.

Figure 2



The substrate fragments. The natural substrate of TS, dUMP (**1**), can be subdivided into several fragments, of which various combinations are commercially available. These include 2′-deoxyuridine (dUrd; **2**), 2′,5′-dideoxyuridine (ddUrd; **3**), uridine (Urd; **4**), phosphoribose (PR; **5**), and phosphate (PO_4^{2-}; **6**).

Substrate influence on protein structure

By comparing the structure of the apoenzyme with that of TS in complex with the natural substrate, dUMP, we can elucidate the influence of the ligand on the protein structure. As the apoenzyme crystals are grown from 20 mM KPO_4 buffer, the phosphate-binding site is fully occupied in the apoTS crystal structure. Thus, the comparison is, in effect, between the complexes of TS with the bound substrate fragments PO_4^{2-} (**6**) and dUMP (**1**). Globally, there are minimal effects on the structure of TS following the binding of dUMP. The root mean square (rms) deviation for Cα atoms between the two complexes is 0.18 Å (Table 1), well within the experimental error. Within the active site, however, there are a number of rearrangements by the specificity framework of the protein to accommodate the substrate molecule. Two well ordered water molecules are displaced by the dUMP; water 403 occupied a site in the PO_4^{2-} bound structure very near C6 of dUMP,

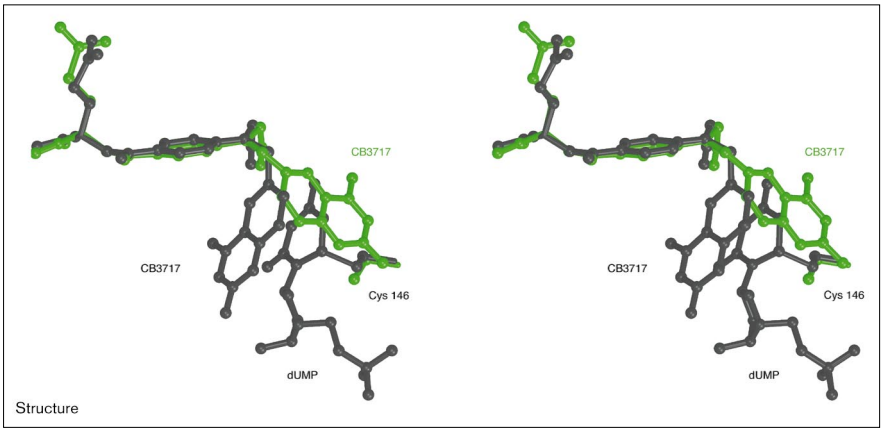
Table 1

Structural comparisons.

	apo	+ dUMP	+ dUrd	+ dUMP + CB	+ dUrd + CB	+ ddUrd + CB	+ CB
3TMS							
Cα	0.19 (7.9)						
main	0.19 (7.9)						
all	0.53 (9.9)						
2TSC							
Cα				0.22 (5.4)			
main				0.22 (5.4)			
all				0.78 (6.7)			
apoTS							
Cα		0.18 (4.9)	0.18 (3.3)				
main	–	0.17 (4.9)	0.18 (3.3)				
all		0.78 (6.3)	0.63 (4.9)				
TS + dUMP + CB3717							
Cα					0.24 (5.6)	0.25 (5.7)	0.23 (7.0)
main				–	0.27 (5.8)	0.25 (5.7)	0.21 (7.4)
all					0.73 (6.5)	0.67 (6.6)	0.56 (8.1)

The values given are root mean square deviations (rmsds) on atomic coordinates (x,y,z); values in parentheses are rmsds on the differences in B factors (rms ΔB). All values are in $Å^2$ and all superpositions were performed using the program lsqman [35]. CB, CB3717; Cα, α carbons only; main, backbone atoms (C, N, Cα); all, all protein atoms. 3TMS and 2TSC are the PDB entry codes for prior determinations of the apoTS and TS + dUMP + CB3717 complexes, respectively.

Figure 3



Stereoview superposition of the binary complex of *E. coli* TS with CB3717 (1AN5; green) and the ternary complex of TS with CB3717 and dUMP (1KCE; dark grey). In the binary complex of CB3717 with TS, the cofactor analogue adopts the 'precompetent' binding mode which is accessible in the absence of substrate. In this figure, and figures 4–6, the 'reference' ligands – dUMP and CB3717 – are depicted in dark grey, while each of the structures of fragment complexes are depicted in a unique colour. The catalytic cysteine of the enzyme is depicted in the colour corresponding to the appropriate complex.

and water 365 is displaced from a point occupied by C4 of the substrate. The sidechain of Asn177 rotates by 180° about the Cα–Cβ bond and Cγ is translated by 1.2 Å to form dual hydrogen bonds with N3 and O4 of dUMP. The sidechain of Gln165 also flips, and Cδ is translated by 0.8 Å to preserve the hydrogen-bonding network. Trp83 and Val143 are also seen to reorient into rotomers 180° opposed to those found in the PO_4^{2-} only structure. There is very little movement in the residues that form hydrogen bonds with the ribose moiety; Tyr209, His207 and Asp169 all move 0.35 Å. This region of the active site appears to be preorganized for substrate binding.

Cofactor influence on protein structure

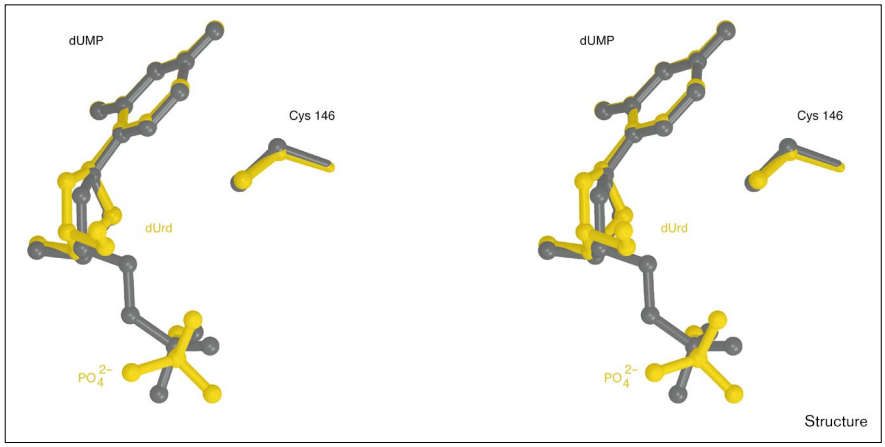
The apoenzyme structure can also be compared to the crystal structure of TS in complex with the cofactor analogue CB3717 alone as a means of discerning the influence of the cofactor analogue on the conformation of the protein. The structure of TS in complex with CB3717 supports previous observations [4,7–9] that it is the cofactor that triggers 'closure' of the active site, principally through the movement of the C terminus towards the reactive center. The cofactor facilitates this through a number of contacts, both specific and hydrophobic. The quinazoline moiety participates in a hydrogen-bond network bridging the C terminus to the arginine pocket that binds the phosphate of the substrate. In addition, the aromatic des-amino ring of the quinazoline and the benzene ring of the *para*-amino benzoic acid (PABA) moieties form favourable hydrophobic-packing interactions with a large portion of the active site, drawing distal walls together and excluding nonspecific waters. In contrast to an earlier study [4], our analysis of the crystal structure of TS complexed with CB3717 in the absence of substrate shows a distinctly different binding mode for the cofactor analogue (Figure 3). Our analysis of this complex shows that what appeared to be disordered binding in the previously published structure of this complex is — at least in this redetermination — a well ordered binding mode which is fully occupied in both monomers of the crystallographically unique dimer. The cofactor adopts an orientation very similar to that we recently described in the second monomer of the Glu58→Gln mutant crystal structure [10]. The quinazoline is rotated about the C9–N10 bond angle by ~90°, bringing N2 to a separation of 3.0 Å from the reactive thiol at Cys146. In addition, the N1–C4 portions of the quinazoline pass essentially through the space occupied by C6–C5 of the pyrimidine ring in the ternary complex with CB3717 and dUMP. The volume occupied by the quinazoline moiety in its 'catalytically competent' orientation is, in this structure, filled with three well ordered water molecules: Wat767, Wat781 and Wat789. In addition to the close interaction with the active-site cysteine, CB3717 makes four additional hydrogen bonds which are unique to this conformation. These interactions are with Ala144 (N2′→O; 2.6 Å), Tyr94 (OA4′→Oη; 2.8 Å), His147 (OA4′→Nε2; 3.2 Å), and the conserved water that anchors O4 of dUMP (OA4′→Wat771; 2.9 Å). This conformation of the cofactor analogue may represent a 'pre-ordered' or 'pre-catalytically competent' orientation of the cofactor. In addition, the very close interaction between the free amine at the tip of the quinazoline and the active-site cysteine immediately presents itself as an intriguing new avenue of inquiry for rational drug design against TS. Modification of N2 from a primary amine to a functional group that can covalently interact with the reactive-site cysteine would bring the mechanism-based advantages of substrate analogues, for example 5-fluorouracil (5FU) [11,12], together with the cell accessibility advantages of a cofactor analogue, for example CB3717 [13–18], BW1843U89 [13,19,20] and Tomudex [21–23].

Effect of the cofactor on dUMP binding

A number of differences are induced in the protein structure on cofactor binding; most of these have been

Figure 4

Stereoview superposition of the binary complexes of *E. coli* TS with dUrd (1BDU; yellow) and TS with dUMP (1BID; dark grey). The absence of the covalent tether between dUrd and the PO_4^{2-} allows movement of the ribose moiety and the phosphate ion away from one another relative to their positions in the dUMP complex.



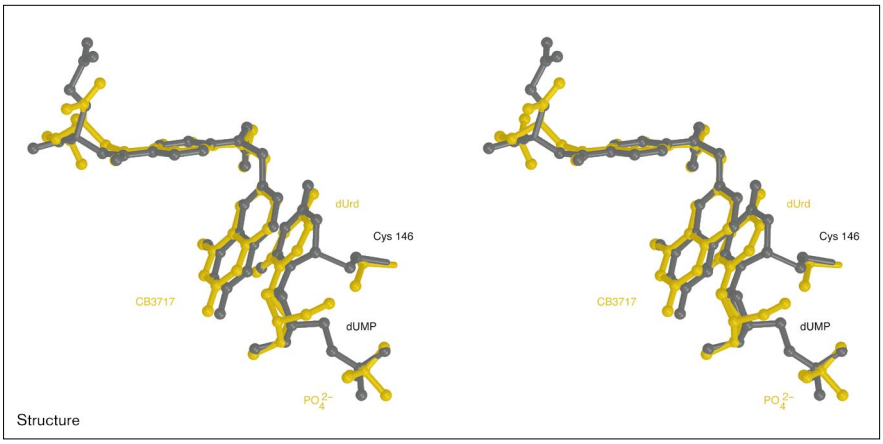
discussed previously [4,7,8]. Principally, they include a 'closure' of the active site through a concerted inward movement of the C terminus toward the cofactor. A comparison of the binary TS–dUMP complex with the ternary complex formed between TS, dUMP and CB3717 demonstrates the effect that the cofactor exerts on the binding mode of the substrate. By aligning the proteins through a least-squares fit of the Cα atoms which deviate by less than 1 Å from the aligned positions (thereby selecting a best-fitting subset of the protein and neglecting regions of large conformational change; in this case the C-terminal four residues) [24], a direct comparison of the active sites with substrate and both with and without the presence of cofactor was made. The most immediately obvious difference is that in the ternary complex, dUMP forms a covalent adduct with Sγ of Cys146. This is a well known initial step in the reaction mechanism [1], and is trapped by the cofactor analogue CB3717 [25]. In addition, it is clear that the presence of the cofactor analogue induces some increase in conformational energy in the substrate by pressing the ribose and pyrimidine moieties closer towards the reactive thiol while the phosphate remains rigidly in place, anchored by the four arginines of the phosphate-binding pocket. On average, the pyrimidine is moved ~0.9 Å towards Cys146, while the ribose moves ~0.7 Å. In particular, C6 of the pyrimidine is moved 1.2 Å, and it has been postulated (reviewed in [26]) that this enhanced proximity of the pyrimidine to the reactive thiol initiates the TS reaction. Binding studies have also shown [27] that the addition of CB3717 enhances the affinity of the substrate by around threefold relative to binary complexes. This increase in affinity is reflected in the degree of order in the crystal structures of dUMP in complex with TS both with and without bound antifolate inhibitor CB3717: after normalization based on the average B factor of the protein, the average difference in thermal parameters for the substrate (<ΔB>) is 21.3 Å^2 (relative to an average ΔB of zero for the normalised protein B factors) suggesting that the increase in affinity is a result of a decrease in the local entropy of the ligand. Thus, the addition of CB3717 has principally contributed to an ordering of dUMP binding, as well as to a localisation of the substrate in close proximity to the reactive thiol.

'Loss' of the glycosidic bond

2′-Deoxyuridine (dUrd) differs chemically from the natural substrate, dUMP, by the absence of the phosphate moiety (see Figure 1). The 'loss' of this phosphate group reduces the binding of dUrd (K_m = 0.2 mM [28,29]) to TS by 500-fold relative to dUMP. We have measured the propensity of dUrd to inhibit the *Escherichia coli* TS reaction and found no measurable diminution of enzyme activity, even in the presence of 20 mM dUrd. When the natural substrate, dUMP, is substituted in the binary crystalline complex by 2′-deoxyuridine (dUrd) and a phosphate ion, PO_4^{2-} (**6**), there are minimal differences found in the mode of binding (Figure 4; Table 1). There is a small shift (~0.2 Å) in the center of mass of the dUrd molecule away from the phosphate-binding site and toward Asn177 which coordinates the pyrimidine ring. There is a corresponding movement of the Asn177 sidechain to track this movement, thereby maintaining the double hydrogen bonds between Asn177 and N3 and O4 of the pyrimidine.

It is in the ternary complexes of dUMP and dUrd with TS and CB3717, that the most distinct differences are found between the binding aspects of the substrate fragments (Figure 5). In the binary complexes, dUMP and dUrd bind in a very similar manner, but the presence of the cofactor analogue demonstrates the very different interactions with the enzyme that these substrate fragments make. Most prominently, whereas the presence of CB3717 promotes the covalent addition of dUMP to Sγ of

Figure 5



Stereoview superposition of the ternary complexes of *E. coli* TS with CB3717 and dUrd (1TDU; yellow) and with CB3717 and dUMP (1KCE; dark grey). These complexes differ from those depicted in Figure 4 by the additional presence of the cofactor analogue CB3717. Here we see that whereas CB3717 facilitates the covalent attachment of dUMP to Cys146, there is no crystallographically detectable attachment between dUrd and the enzyme.

Cys146, this Michael adduct is not formed in the ternary complex of CB3717 with dUrd. The effect of CB3717 on the binding mode of dUrd is similar to the difference found between the binary dUMP complex and the ternary TS–CB3717–dUMP complex, where the presence of the cofactor analogue has pressed the pyrimidine ring of dUMP closer to Cys146. In the TS–CB3717–dUrd ternary complex, the pyrimidine ring is moved ~0.9 Å in the direction of Cys146 relative to the binary dUrd complex; however, there is no covalent attachment made between C6 of dUrd and the Sγ of Cys146. The presence of a covalent attachment is very clear in the TS–dUMP–CB3717 ternary complex: C6 of dUMP and Sγ of Cys146 are separated by ~1.8 Å, and the pyrimidine ring of dUMP can clearly be seen to be nonplanar. Each of these features was determined from difference electron-density maps from which dUMP/dUrd and Cys146 have been omitted. In the ternary TS–dUrd–CB3717 complex, difference electron density unambiguously demonstrated the planarity of the pyrimidine, as well as the much larger separation between C6 and Sγ which, after refinement, was found to be 3.8 Å. In addition, the cysteine could be seen to have moved ~1 Å away from the pyrimidine-binding site. Without the constraints of the phosphate fragment, dUrd is found to undergo a concerted rotation away from both the phosphate-binding site and the reactive thiol. The side of dUrd that anchors the molecule to the protein, encompassing C2–N3 and CR2–CR3, moves less than 0.1 Å from the same atomic positions found in the ternary complex with dUMP. The opposite side of the molecule, however, encompassing C5–C6 and OR1–CR1, rotates away from Cys146 and the phosphate-binding site by an average of 0.5 Å. The inability of dUrd to covalently attach to TS in the presence of the cofactor analogue, indicates that the phosphate moiety and its interaction with the remainder of the substrate have a fundamental role in the initial stages of the TS reaction.

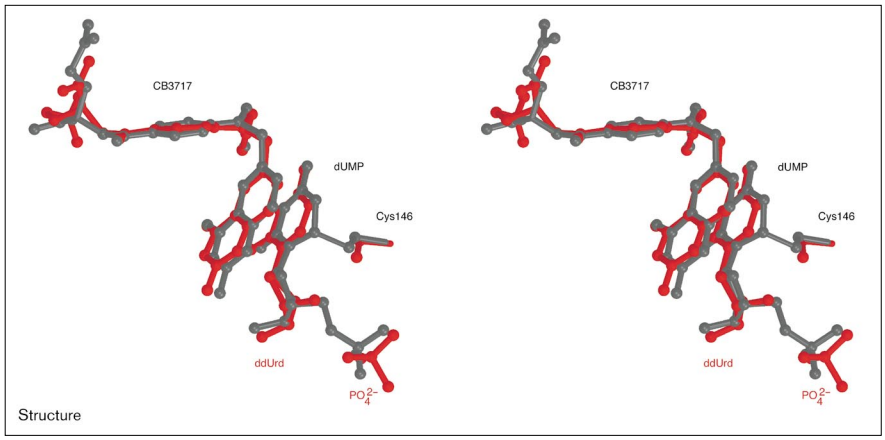
Effect of the 5′-terminal hydroxyl

To investigate the role of the 5′-hydroxyl group included on dUrd (and essentially duplicated in the phosphate ion), we also cocrystallised and determined the structures of binary and ternary complexes of TS with *2′*,5′-dideoxyuridine (ddUrd). In both cases, the binding mode of the ddUrd moiety (Figure 6) is nearly identical to that observed for dUrd. ddUrd displays the same 'released' and relaxed behaviour relative to the binding mode of dUMP as described above for dUrd. Both binary and ternary complexes of ddUrd with TS do, however, show larger B factors for the ddUrd after normalisation of the protein B factors against the dUrd complexes. This would seem to indicate that, although the 5′-hydroxyl does not form any apparent hydrogen bonds with the protein or with the crystallographically ordered solvent, it does seem to play some part in reducing the local entropy of the binding mode.

If the difference between dUMP and dUrd/ddUrd can be thought of as the 'loss' of the glycosidic bond, then the complexes can be compared as 'before' and 'after' endpoints encompassing the breakage of this bond. The complexes of TS with dUrd or ddUrd and phosphate show that the release of the anchoring aspects of the phosphate moiety allows global movement of dUrd/ddUrd away from the phosphate-binding site as well as a rotation of the 5′-terminal hydroxyl group away from the phosphate ion (in the case of dUrd). As dUrd/ddUrd is no longer anchored into the phosphate-binding site, there is a rotational shift of the ribose ring away from the phosphate-binding site, presumably relieving the substantial amount of strain that is induced into dUMP by its multiple site attachments. We propose that it is, in part, this stereochemical strain induced into dUMP by the anchoring of the phosphate and the pyrimidine to the enzyme, and enhanced by the presence of the cofactor, which activates

Figure 6

Stereoview superposition of the ternary complexes of *E. coli* TS with CB3717 and ddUrd (1DDU; red) and with CB3717 and dUMP (1KCE; dark grey).



C6 of the pyrimidine ring, making this carbon susceptible to attack by the active-site thiol.

'Loss' of the ribose

Binary and ternary complexes were also obtained containing both of the substrate fragments uridine (Urd) and phosphate (PO_4^{2-}). These fragments correspond to all of the components of dUMP binding with the ribose sugar ring removed. In a serial progression from dUMP→dUrd→ddUrd→Urd, the removal of the ribose demonstrates a dramatic effect on the coherence of binding to TS. In both the binary and ternary complexes with Urd, the pyrimidine was found to be very poorly ordered in the active site. The principal orienting interactions with Asn177 are maintained; however, the ligand is librationally disordered. Relative to the complexes with dUrd and ddUrd, it can be inferred that the principal contribution of the ribose moiety to ligand binding is to enhance the binding affinity and reduce the local entropy of the substrate within the active site. This is probably accomplished both through the conformational inflexibility of the five-membered ring and the stereospecificity of the hydrogen bonds formed between O3′ and Tyr209 and His207.

Nonspecific binding of phosphoribose

Attempts were also made to form binary and ternary complexes with phosphoribose (PR). Cocrystallisations were set up in the same manner as for each of the other complexes, and data were collected. Difference electron-density maps, however, did not show readily interpretable electron density for this substrate fragment. The ΔF_o ($F_{PR}–F_{apo}$ and $F_{CB,PR}–F_{CB}$) electron-density maps show a very strong peak for the phosphate ion which is elongated in the direction of the ribose moiety; however, there is no clearly interpretable orientation in which the ribose can be modelled. Similar maps calculated for crystals containing only PO_4^{2-} show very spherical/tetrahedral density at the phosphate-binding site. The comma-shaped elongation of the difference density makes it clear that PR is present at full occupancy; however, the ribose is rotationally disordered about the glycosidic linkage. This result is just as informationally rich as the clearly modelled complexes in telling us what is important about the binding of substrate to the enzyme. The inability of phosphate-tethered ribose to interact in an ordered manner with the active site without the presence of the pyrimidine ring indicates that the pyrimidine is the principal determinant of binding modality for substrate within the active site. In addition, this observation suggests that the role of the ribose is to confer additional specificity, to enhance the strength of the substrate–protein interaction (through multiple hydrogen bonds), and perhaps to serve as a steric bridge between the phosphate-binding site and C6 of the pyrimidine.

Conclusions

This study has shown that analysis of the binding of ligands to enzymes by considering the binding modes and affinities of fragments of the native substrate can reveal insights into the requirements for ligand binding. It can be postulated from these observations that the binding of ligand fragments may conform to the principles of additivity, enabling new in-roads into the design of enzyme inhibitors and therapeutics. This study of protein–ligand interactions via the interactions of the enzyme with ligand fragments is complementary to approaches utilising site-directed mutagenesis of the enzyme. In the case of TS, we have found that the binding orientation of dUMP is principally directed by the pyrimidine ring. We have also found that the ribose sugar moiety contributes a certain degree of specificity, as other sugars are not tolerated, and also greatly enhances the spatial localisation of substrate within the active site. In addition, we have found that the glycosidic linkage is critical to the precise localisation and activation of the substrate for initiation of the enzymatic

reaction. This reduction in local entropy may be crucial to the turnover rate of enzymatic activity.

Biological implications

Throughout nature, enzymes have responded to evolutionary pressures by evolving the ability to recognise their target substrates with extraordinary selectivity and specificity. For example, *Lactobacillus casei* thymidylate synthase (TS) preferentially binds dUMP over dCMP by approximately 500-fold [30] via a mechanism that has recently been exploited to produce a re-engineered TS with a tenfold reversal in ligand-recognition affinity [31]. The complex interactions which have evolved between ligand and enzyme complicate the process of both understanding the exceptional chemistry performed by enzymes, as well as our ability to readily exploit the active-site features that must be targeted in the design of enzymatic inhibitors. Optimisation of this process is of fundamental importance to streamlining the design of medicinal therapeutics if we are to move lead compounds to clinical candidates in a timely manner.

Structural as well as kinetic studies of the interactions between enzymes and their substrates have generally found that multiple interactions are often generated around the available 'hooks' presented by the substrate. Whether fragments of the substrate —probably never available to the evolving enzyme —are bound in the same manner as the parent substrate addresses the fundamental basis of specificity. An understanding of the relative contributions of individual portions of ligand molecules to the enzyme-binding interaction may offer considerable insight into the principles of substrate recognition.

In all cells, TS provides the last step in the only synthetic pathway for the production of thymidine. Because rapidly dividing cells, such as those in the cancerous state, have a particularly high demand for DNA production and thus for thymidine, TS has become an attractive target for anticancer drug design. Much work has focused on TS as a drug design target; however, to date, only the pro-drug 5-fluorouracil (5FU) [11,12] is widely approved as a TS-inhibiting anticancer therapy. 5FU is a mechanism-based inhibitor, meaning that it takes advantage of the natural ligand-binding affinities and chemistry of TS to 'trap' the enzyme in a complex that can no longer produce products. In this study, we have sought to break apart the natural binding affinity of the substrate, dUMP, into its various modular components. An understanding of which components of the substrate are crucial to binding and to the chemical mechanism will lead to better rational design of anticancer drugs directed against this enzyme. Towards this end, we have examined the crystal structures of TS in complex with rational chemical subsets of dUMP. On the basis of these structures, we can deduce the relative importance of each fragment to the interactions made between the natural substrate and the protein. This approach complements the understanding of ligand–protein interactions obtained by site-directed mutagenesis, wherein the protein is modulated through amino acid changes and the resulting effects on the binding of substrate are studied.

Materials and methods

Substrate fragments

All the substrate fragments were purchased from Sigma. They include the intact substrate 2′,5′-deoxyuridine monophosphate (dUMP, **1** in Figure 1), a fragment incorporating all of dUMP minus the phosphate, 2′-deoxyuridine (dUrd, **2**); a fragment which is also missing the ribose/phosphate 'linkage' oxygen, 2′,5′-dideoxyuridine (ddUrd, **3**); a fragment consisting only of the pyrimidine ring, uracil (Urd, **4**); a fragment lacking the pyrimidine, phosphoribose (PR, **5**); and a fragment consisting only of the phosphate moiety (PO_4^{2-}, **6**).

Cocrystallisation

E. coli TS was expressed and purified as described previously [10]. Crystals of each of the complexes were formed by cocrystallisation of TS with the ligands. 'Binary' crystallisation experiments (those not containing the cofactor analogue, CB3717) were conducted in hanging drops containing ~4.2 mg/ml *E. coli* TS, 0.38 mM dUMP, dUrd, ddUrd, Urd or PR, 3.8 mM DTT, and 1.2 M $(NH_4)_2SO_4$, at pH 7.8 (20 mM KPO_4) suspended over a well solution containing 2.4 M $(NH_4)_2SO_4$ and 1.0 mM DTT. 'Ternary' crystallisation experiments (those including CB3717) were conducted in hanging drops containing ~4.2 mg/ml *E. coli* TS, 0.38 mM dUMP, dUrd, ddUrd, Urd or PR, 3.8 mM DTT, 1.0 mM CB3717 and 1.2 M $(NH_4)_2SO_4$, pH 7.8 (20 mM KPO_4) over a well solution containing 2.4 M $(NH_4)_2SO_4$ and 1.0 mM DTT. The final volume of the droplets was 5 µl. Cubes and icosahedra (binary complexes) or hexagonal rods (ternary complexes) grew within 3 days. The largest binary complex crystals typically grew to 350–400 μm^3, while ternary crystals were typically $250 \times 250 \times 600$ µm. Binary crystals belong to the cubic space group $I2_13$, with a = 133.2 Å, and ternary crystals belong to space group $P6_3$, with a = 127.35 Å and c = 68.16 Å. The cubic form contains one *E. coli* TS monomer per asymmetric unit, whereas the hexagonal form contains a full dimer in the asymmetric unit. In a small percentage of crystallisations, ternary complex crystals are found in the trigonal spacegroup, $P3_121$ (a = 71.96 Å, c = 115.1 Å).

X-ray data collection

All X-ray diffraction data were collected on an R-Axis IIc imaging plate with a Rigaku RU-200 rotating anode generator operating at 15 kW (50 mA and 300 kV) fitted with a Cu anode ($\lambda = 1.5418$ Å). The crystal-to-detector distance was 100 mm for the body-centered cubic crystals (apo and 'binary' complexes), and 120 mm for the primitive lattice hexagonal crystals ('ternary' complexes). Exposures of 20 min per 1° oscillation range were used throughout all of the data collections. The diffraction data were indexed, integrated, scaled and merged with the HKL software package [32]. A summary of the data-processing statistics is presented in Table 2.

Structure solution and refinement: binary complexes

For each of the binary complexes, a model consisting of one monomer of *E. coli* TS in the binary complex form [7], excluding all waters and ligands was used to calculate an initial difference electron density synthesis (F_o–F_c) using all data between 15-2.5 Å. dUMP or dUMP fragments were built into the resulting difference density and refined using X-PLOR [33]. Waters were added to the structures by an automated peak picking procedure which selected 3σ or greater peaks from F_o–F_c difference density maps which also satisfied proper criteria for water

Table 2

Crystallographic data statistics*.

	apo	+ dUMP	+ dUrd	+ ddUrd	+ Urd	+ dUMP + CB	+ dUrd + CB	+ ddUrd + CB	+ Urd + CB	+ CB
Resolution (Å)	2.20	2.10	2.10	2.10	2.20	2.00	2.10	2.10	2.50	2.6
Observed reflections†	61,026	92,187	68,944	76,509	32,842	109,083	148,364	55,680	29,322	76,707
Unique reflections	18,520	22,297	21,786	20,597	17,015	37,807	35,043	25,970	15,999	19,013
Completeness (%)	92.7	96.9	95.0	89.7	85.1	90.0	95.2	70.1	72.8	97.4
	(85.7)	(91.5)	(89.2)	(74.5)	(71.7)	(69.3)	(90.2)	(49.8)	(52.8)	(94.3)
R_{sym} (%)	0.070	0.074	0.096	0.098	0.076	0.097	0.064	0.133	0.132	0.109
	(0.273)	(0.319)	(0.366)	(0.340)	(0.207)	(0.261)	(0.278)	(0.252)	(0.262)	(0.300)
R_{merge} (%)	21.3‡					23.4§				
⟨Redundancy⟩	3.5	4.2	3.2	3.7	1.9	2.9	4.2	2.1	1.8	4.0
Refinement resolution (Å)	7–2.2	7–2.1	7–2.1	7–2.1	7–2.5	7–1.95	7–2.1	7–2.1	7–2.5	7–2.6
Refined R factor	0.190	0.179	0.181	0.193	0.160	0.172	0.185	0.162		0.162
R free	0.243	0.217	0.237	0.243	0.221	0.221	0.246	0.256		0.244
No. of modelled waters	99	113	68	0	106	241	153	66		172
Average B (protein) (Å^2)	23.3	26.9	21.1	22.0	20.8	20.9	24.8	22.8		17.44
Average B (substrate) (Å^2)	–	53.2	77.1	72.7	ind#	16.9	35.6	84.9		–
Average B (CB3717) (Å^2)	–	–	–	–	–	27.3	47.1	75.7		39.87
Average B (water) (Å^2)	39.5	40.6	39.0	–	36.6	35.7	33.1	37.0		34.04

*Numbers in parentheses refer to the highest resolution shell; CB, CB3717. †($I/\sigma(I) > 1.0$). ‡versus 3TMS and §versus 2TSC, where 3TMS and 2TSC are the PDB entry codes for previous determinations of the apoTS and ternary TS + dUMP + CB3717 structures, respectively. #ind, indeterminate.

molecule positions – including at least one proper hydrogen-bonding partner (either protein, ligand or bridging waters) and good stereochemistry. Once all of the significant difference density was accounted for, refinement was extended to the ultimate resolution limit of the structure being considered. All of the structures were examined for adherence to good stereochemical parameters without over-refinement (Table 3).

Structure solution and refinement: ternary complexes

For each of the ternary complexes, a model of a complete dimer of *E. coli* TS in the ternary complex form [8], excluding all waters and ligands was used to calculate an initial difference electron density synthesis (F_o–F_c) using all data between 15–2.5 Å. CB3717 was built into the resulting difference density first, using the density visualisation program, CHAIN [34], and refined against the diffraction data using the program X-PLOR [33]; subsequent difference density maps were used to place the dUMP or dUMP fragments which were built into this difference density and refined. Waters were added to the structures by an automated peak picking procedure which selected 3σ or greater peaks from F_o–F_c difference density maps which also satisfied proper criteria for water molecule positions – including at least one proper hydrogen-bonding partner (either protein, ligand or bridging waters) and good stereochemistry. Once all of the significant difference density was accounted for, refinement was extended to the ultimate resolution limit of the structure being considered, and each of the structures was examined for adherence to good stereochemical parameters without over-refinement (Table 3).

Accession numbers

The coordinates of the structure of TS and the various complexes reported here have been deposited with the Brookhaven Protein Data Bank. Entry codes: apoTS, 1TJS; binary TS + dUMP, 1BID; binary TS + dUrd + PO_4^{2-}, 1BDU; binary TS + ddUrd + PO_4^{2-}, 1AOB; ternary TS + CB3717 + dUMP, 1KCE; ternary TS + CB3717 + dUrd + PO_4^{2-}, 1TDU; ternary TS + CB3717 + ddUrd + PO_4^{2-}, 1DDU; ternary TS + CB3717 + PO_4^{2-}, 1AN5; crystallographically symmetric ternary TS + CB3717 + dUMP, 1TRG.

Table 3

Stereochemical summaries.

Accession No.	Ramachandran (%)*	Bond lengths (%)†	Bond angles (%)†	Planar groups (%)†	G factors‡
3TMS	91.3, 8.7, 0.0, 0.0	87.7, 12.3	62.7, 37.3	88.9, 11.1	–0.29, –0.66, –0.39
2TSC	89.8,10.0, 0.0, 0.2	73.0, 27.0	43.8, 56.2	94.3, 5.7	–0.10, –2.20, –0.82
1TJS	89.6, 10.4, 0.0, 0.0	100.0, 0.0	83.7, 16.3	86.3, 13.7	0.05, 0.41, 0.20
1KCE	92.8, 7.2, 0.0, 0.0	100.0, 0.0	86.6, 13.4	88.6, 11.4	0.21, 0.47, 0.32
1BDU	92.2, 7.8, 0.0, 0.0	98.0, 2.0	73.7, 26.3	89.3, 10.7	–0.14, –0.19, –0.13
1TDU	89.8, 9.8, 0.4, 0.0	97.3, 2.7	71.5, 28.5	88.6, 11.4	–0.19, –0.13, –0.13
1AOB	91.7, 8.3, 0.0, 0.0	95.3, 4.7	74.8, 25.2	96.4, 3.6	–0.10, –0.13, –0.08
1DDU	87.8, 12.0, 0.2, 0.0	91.7, 8.3	67.7, 32.3	92.9, 7.1	–0.23, –0.16, –0.20
1AN5	89.1, 10.9, 0.0, 0.0	100.0, 0.0	86.5, 13.5	82.6, 17.4	0.15, 0.48, 0.29

*The percentage of residues falling within core, allowed, generously allowed, and disallowed zones [36]. †The percentage of residues falling within generally observed limits, and the percentage flagged as unusual . ‡G factors for dihedral angles, covalent bonds and the overall combined value [37].

Acknowledgements

This work was supported by a grant from the National Institutes of Health to RMS (R01 CA 63081). TJS was supported by a post-doctoral fellowship from the American Cancer Society. CRS was supported by a post-doctoral fellowship from the National Institutes of Health (AI09211). We thank Linda S Brinen for expert assistance in sample preparation and data collection.

References

1. Carreras, C.W. & Santi, D.V. (1995). The catalytic mechanism and structure of thymidylate synthase. *Annu. Rev. Biochem.* **64**, 721-762.
2. Hardy, L.W., *et al.*, & Stroud, R.M. (1987). Atomic structure of thymidylate synthase: target for rational drug design. *Science* **235**, 448-455.
3. Finer-Moore, J.S., Montfort, W.R. & Stroud, R.M. (1990). Pairwise specificity and sequential binding in enzyme catalysis: thymidylate synthase. *Biochemistry* **29**, 6977-6986.
4. Kamb, A., Finer-Moore, J.S. & Stroud, R.M. (1992). Cofactor triggers the conformational change in thymidylate synthase: implications for an ordered binding mechanism. *Biochemistry* **31**, 12876-12884.
5. Finer-Moore, J., *et al.*, & Stroud, R.M. (1993). Refined structures of substrate-bound and phosphate-bound thymidylate synthase from *Lactobacillus casei*. *J. Mol. Biol.* **232**, 1101-1116.
6. Stroud, R.M. & Finer-Moore, J.S. (1993). Stereochemistry of a multistep/bipartite methyl transfer reaction: thymidylate synthase. *FASEB J.* **7**, 671-677.
7. Perry, K.M., *et al.*, & Stroud, R.M. (1990). Plastic adaptation toward mutations in proteins: structural comparison of thymidylate synthases. *Proteins* **8**, 315-333.
8. Hartman, K.-R. & Heidelberger, C. (1958). Studies on fluorinated pyrimidines: XIII. Inhibition of thymidylate synthase. *J. Biol. Chem.* **236**, 3006-3013.
9. Heidelberger, C. (1965). Fluorinated pyrimidines. *Prog. Nucleic Acid Res. Mol. Biol.* **4**, 1-50.
10. Montfort, W.R., *et al.*, & Stroud, R.M. (1990). Structure, multiple site binding, and segmental accommodation in thymidylate synthase on binding dUMP and an anti-folate. *Biochemistry* **29**, 6964-6977. Erratum. (1990). *Biochemistry* **29**, 10864.
11. Carreras, C.W., Naber, N., Cooke, R. & Santi, D.V. (1994). A C-terminal conformational equilibrium in thymidylate synthase observed by electron paramagnetic resonance spectroscopy. *Biochemistry* **33**, 2071-2077.
12. Sage, C.R., Rutenber, E.E., Stout, T.J. & Stroud, R.M. (1996). An essential role for water in an enzyme reaction mechanism: the crystal structure of the thymidylate synthase mutant E580. *Biochemistry* **35**, 16270-16281.
13. Westerhof, G.R., *et al.*, & Pinedo, H. M. (1995). Carrier- and receptor-mediated transport of folate antagonists targeting folate-dependent enzymes: correlates of molecular-structure and biological activity. *Mol. Pharmacol.* **48**, 459-471.
14. Westerhof, G.R., *et al.*, & Schornagel, J.H. (1991). Membrane transport of natural folates and antifolate compounds in murine L1210 leukemia cells: role of carrier- and receptor-mediated transport systems. *Cancer Res.* **51**, 5507-5513.
15. Newell, D.R., Alison, D.L., *et al.*, & Jones, T.R. (1986). Pharmacokinetics of the thymidylate synthase inhibitor N^{10}-propargyl-5,8-dideazafolic acid (CB3717) in the mouse. *Cancer Treat. Rep.* **70**, 971-979.
16. Jansen, G., *et al.*, & Jackman, A.L. (1990). Multiple membrane transport systems for the uptake of folate-based thymidylate synthase inhibitors. *Cancer Res.* **50**, 7544-7548.
17. Jackson, R.C., Jackman, A.L. & Calvert, A.H. (1983). Biochemical effects of a quinazoline inhibitor of thymidylate synthase, CB3717, on human lymphoblastoid cells. *Biochem. Pharmacol.* **32**, 3783-3790.
18. Alison, D.L., *et al.*, & Calvert, A.H. (1985). The clinical pharmacokinetics of the novel antifolate N^{10}-propargyl-5,8-dideazafolic acid (CB3717). *Cancer Chemother. Pharmacol.* **14**, 265-271.
19. Duch, D.S., *et al.*, & Wilson, H.R. (1993). Biochemical and cellular pharmacology of 1843U89, a novel benzoquinazoline inhibitor of thymidylate synthase. *Cancer Res.* **53**, 810-818.
20. Hanlon, M.H. & Ferone, R. (1996). *In vitro* uptake, anabolism, and cellular retention of 1843U89 and other benzoquinazoline inhibitors of thymidylate synthase. *Cancer Res.* **56**, 3301-3306.
21. Gibson, W., *et al.*, & Jackman, A.L. (1993). The measurement of polyglutamate metabolites of the thymidylate synthase inhibitor, ICI D1694, in mouse and human cultured cells. *Biochem. Pharmacol.* **45**, 863-869.
22. Jodrell, D.I., Newell, D.R., Gibson, W., Hughes, L.R. & Calvert, A.H. (1991). The pharmacokinetics of the quinazoline antifolate ICI D 1694 in mice and rats. *Cancer Chemother. Pharmacol.* **28**, 331-338.
23. Ward, W.H., Kimbell, R. & Jackman, A.L. (1992). Kinetic characteristics of ICI D1694: a quinazoline antifolate which inhibits thymidylate synthase. *Biochem. Pharmacol.* **43**, 2029-2031.
24. Kleywegt, G.J. & Jones, T.A. (1994). A super position *ESF/CCP4 Newslett.* **31**, 9-14.
25. Pogolotti, A.L., Jr., Danenberg, P.V. & Santi, D.V. (1986). Kinetics and mechanism of interaction of 10-propargyl-5,8-dideazafolate with thymidylate synthase. *J. Med. Chem.* **29**, 478-482.
26. Santi, D.V. & Danenberg, P.V. (1984). In *Folates and Pterins*, *Vol. I.* (Blakely, R.L. & Bencovic, S.J., eds), pp. 345-399, Wiley, New York.
27. Dev, I.K., Yates, B.B., Hanlon, M., Ferone, R. & Dallas, W.S. (1992). Mode of binding of the folate analogs PDDF and BW1843U89 to human thymidylate synthase. *Abstract from the Meeting of the American Association for Cancer Research,* 2429.
28. Lockshin, A. & Danenberg, P.V. (1981). Biochemical factors affecting the tightness of 5-fluorodeoxyuridylate binding to human thymidylate synthetase. *Biochem. Pharmacol.* **30**, 247-257.
29. Reyes, P. & Heidelberger, C. (1965). Fluorinated pyrimidines. XXVI. Mammalian thymidylate synthetase: its mechanism of action and inhibition by fluorinated nucleotides. *Mol. Pharmacol.* **1**, 14-30.
30. Liu, L. & Santi, D.V. (1992). Mutation of asparagine 229 to aspartate in thymidylate synthase converts the enzyme to a deoxycytidylate methylase. *Biochemistry* **31**, 5100-5104.
31. Agarwalla, S., *et al.*, & Santi, D.V. (1997). A Novel dCMP methylase by engineering thymidylate synthase. *Biochemistry* **36**, 15909-15917.
32. Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend: 'Data Collection and Processing'.* (Sawyer, L., Isaacs, N. & Bailey, S., eds), SERC Daresbury Laboratory, UK.
33. Brünger, A.T. (1992). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR.* Yale University Press, New Haven, CT.
34. Sack, J.S. (1988). CHAIN – a crystallographic modelling program. *J. Mol. Graph.* **6**, 244-245.
35. Kleywegt, G.J. & Jones, T.A. (1995). Where freedom is given, liberties are taken. *Structure* **3**, 535-540.
36. Kleywegt, G.J. & Jones, T.A. (1996). Phi/Psi-chology: Ramachandran revisited. *Structure* **4**, 1395-1400.
37. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.